

# Analyse statistique des séries chronologiques

M1. MAS, M. BOUTAHAR, M. CARENZI

## Definition

On appelle **processus aléatoire du second ordre** à temps discret, toute suite réelle  $Z_t : \Omega \longrightarrow \mathbb{R}$  définie sur un espace de probabilité  $(\Omega, \mathcal{F}, P)$  à valeurs dans  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ , indexée par  $t \in \mathbb{N}$  ou  $\mathbb{Z}$  telle que:

$$E(|Z_t|^2) < \infty \text{ pour tout } t.$$

Pour de tel processus, on peut définir les deux fonctions suivantes:

- i) Moyenne:  $\mu : \mathbb{Z} \rightarrow \mathbb{R}$ ,  $\mu(t) = E(Z_t)$ ,  $\forall t \in \mathbb{Z}$ .
  - ii) Auto-covariance:  $K : \mathbb{Z} \times \mathbb{Z} \rightarrow \mathbb{R}$ ,
- $$K(t, s) = E((Z_t - \mu(t))(Z_s - \mu(s))), \forall t, s \in \mathbb{Z}.$$

# Contenu

## 1 Processus stationnaires au second ordre

- Processus strictement stationnaire
- Processus faiblement stationnaire

## 2 Processus linéaires

- Test KPSS de stationnarité
- Processus ARMA

## 3 Modélisation d'une série chronologique par un processus ARMA

- Identification du processus ARMA

## Definition

Soit  $Z$  un processus; on dit que  $Z$  est strictement (ou fortement) stationnaire si pour toute suite finie d'instants  $t_1, \dots, t_k$  élément de  $\mathbb{Z}$  et tout entier  $r \in \mathbb{Z}$ , les lois jointes de  $(Z_{t_1}, \dots, Z_{t_k})$  et de  $(Z_{r+t_1}, \dots, Z_{r+t_k})$  sont les mêmes (lois jointes invariantes par translation dans le temps).

# Contenu

## 1 Processus stationnaires au second ordre

- Processus strictement stationnaire
- Processus faiblement stationnaire

## 2 Processus linéaires

- Test KPSS de stationnarité
- Processus ARMA

## 3 Modélisation d'une série chronologique par un processus ARMA

- Identification du processus ARMA

## Definition

Un processus est stationnaire au second ordre (on dit aussi faiblement stationnaire) si moyenne et auto-covariance sont invariantes par translation dans le temps, i.e.:

$$\begin{cases} E(Z_t) = \mu, \forall t \in \mathbb{Z}, \\ \text{et} \\ E(Z_t - \mu)(Z_s - \mu) = E(Z_{t+r} - \mu)(Z_{s+r} - \mu), \quad \forall (t, s, r) \in \mathbb{Z}^3. \end{cases}$$

où  $\mu$  est une constante et  $K(t, s)$  ne dépend plus que de  $t - s$ , donc il existe une fonction  $\gamma : \mathbb{Z} \rightarrow \mathbb{R}$ , telle que:

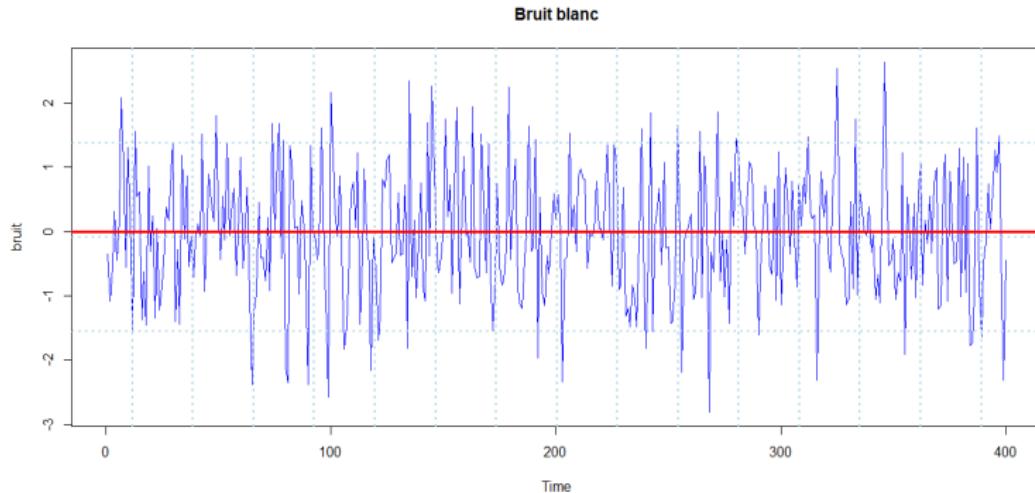
$$K(t, s) = \gamma(t - s).$$

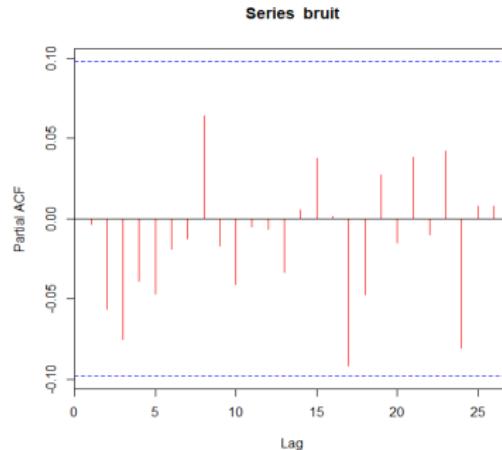
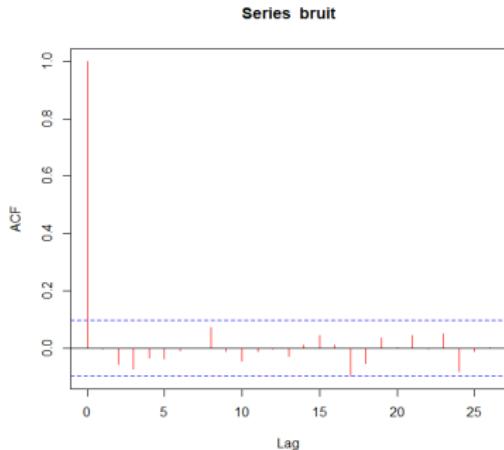
Alors pour un processus faiblement stationnaire, on a

$$\left\{ \begin{array}{l} \text{var}(Z_t) = \gamma(0), \forall t \in \mathbb{Z}, \\ |\gamma(k)| \leq \gamma(0), \forall k \in \mathbb{Z}, \text{ ( par l'inégalité de Cauchy-Schwarz)} \\ \gamma(k) = \gamma(-k), \forall k \in \mathbb{Z}, \\ \rho(j) = \frac{\gamma(j)}{\gamma(0)}, \forall j \in \mathbb{Z}, \text{ fonction d'auto-corrélation ou mémoire} \\ r(j) = \frac{\text{cov}\left(Z_1 - \mathbf{P}_{[z_2, \dots, z_j]}(Z_1), Z_{j+1} - \mathbf{P}_{[z_2, \dots, z_j]}(Z_{j+1})\right)}{\text{var}\left(Z_1 - \mathbf{P}_{[z_2, \dots, z_j]}(Z_1)\right)} \\ \text{fonction d'auto-corrélation partielle.} \end{array} \right.$$

## Bruit blanc

Un bruit blanc est un processus centré  $\varepsilon = (\varepsilon_t)_{t \in \mathbb{Z}}$  tel que  $E(\varepsilon_t \varepsilon_s) = \sigma^2 \delta_t^s = \sigma^2$  si  $t=s$ , et 0 sinon.





**Figure 2:** Les corrélations (ACF) et les corrélations partielles (PACF) du bruit blanc.

Les corrélations (ACF) et les corrélations partielles (PACF), représentées dans la figure 10 ne sont pas significatives.

Un bruit blanc est généralement noté  $(\varepsilon_t)_t$ .

### Definition

- Un processus linéaire  $(Z_t)_t$  est un processus qui peut s'écrire comme une combinaison linéaire d'un bruit blanc :

$$Z_t = \mu + \sum_{k=-\infty}^{\infty} \psi_k \varepsilon_{t-k}, \text{ avec } \psi_0 = 1 \text{ et } \psi_k \in \mathbb{R}. \quad (1)$$

- Un processus linéaire causal  $(Z_t)_t$  est un processus qui peut s'écrire comme une combinaison linéaire du présent et du passé d'un bruit blanc :

$$Z_t = \mu + \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k}, \text{ avec } \psi_0 = 1 \text{ et } \psi_k \in \mathbb{R}. \quad (2)$$

# Contenu

## 1 Processus stationnaires au second ordre

- Processus strictement stationnaire
- Processus faiblement stationnaire

## 2 Processus linéaires

- Test KPSS de stationnarité

- Processus ARMA

## 3 Modélisation d'une série chronologique par un processus ARMA

- Identification du processus ARMA

Soit  $(\varepsilon_t)_t$  un bruit blanc fort de variance  $\sigma_\varepsilon^2$ , et soit  $(X_t)_t$  la marche aléatoire associée :

$$\begin{cases} X_0 &= 0 \\ X_t &= X_{t-1} + \varepsilon_t. \end{cases}$$

Soit  $(B_t)_t$  un processus stationnaire centré indépendant de  $(\varepsilon_t)_t$ . On notera  $v'$  la transposée du vecteur  $v$ . Considérons le modèle

$$Z_t = \beta' D_t + X_t + B_t,$$

où  $D_t$  désigne

- soit la fonction constante égale à 1, et dans ce cas  $\beta' = E(Z_t)$ ,
- soit le vecteur colonne  $(1, t)'$  et dans ce cas  $\beta'D_t$  représente une tendance linéaire.

Le test de stationnarité, proposé par Kwiatkowski-Phillips-Schmidt-Shin (KPSS) en 1992 permet de tester:

$$(H_0) : \sigma_{\varepsilon}^2 = 0 \text{ (et donc } X_t = 0) \text{ contre } (H_1) : \sigma_{\varepsilon}^2 \neq 0.$$

- i) Sous  $(H_0)$ , le processus  $(Z_t)_t$  est stationnaire autour de  $\beta' D_t$  car  $Z_t - \beta' D_t = B_t$ .
- ii) Sous  $(H_1)$ , le processus  $(Z_t)_t$  n'est pas stationnaire autour de  $\beta' D_t$   
car  $\text{var}(Z_t - \beta' D_t) = t \sigma_\varepsilon^2 + \sigma_B^2$ .

# Contenu

## 1 Processus stationnaires au second ordre

- Processus strictement stationnaire
- Processus faiblement stationnaire

## 2 Processus linéaires

- Test KPSS de stationnarité
- Processus ARMA

## 3 Modélisation d'une série chronologique par un processus ARMA

- Identification du processus ARMA

Une classe importante de processus linéaires considérée largement dans la modélisation des séries chronologiques est celle des processus ARMA (AutoRegressive Moving Average).

## Definition

On dit que  $(Z_t)_t$  est un processus **autorégressif moyenne mobile d'ordres  $p$  et  $q$** , noté **ARMA(p,q)** s'il existe un bruit blanc faible  $(\varepsilon_t)_t$  de variance  $\sigma^2$  et des nombres réels  $a_1, \dots, a_p, b_1, \dots, b_q$  tels que

$$Z_t - \mu = \sum_{k=1}^p a_k (Z_{t-k} - \mu) + \varepsilon_t + \sum_{k=1}^q b_k \varepsilon_{t-k}. \quad (3)$$

Ainsi, pour un processus ARMA, la valeur actuelle du processus peut s'écrire comme la somme de deux composantes; la première est une combinaison linéaire du passé du processus lui-même, alors que la deuxième composante est une combinaison linéaire d'un bruit blanc. En particulier:

## Definition

- $(Z_t)_t$  est dit **autorégressif d'ordre  $p$** , noté **AR(p)**, si

$$\begin{aligned} Z_t - \mu &= a_1(Z_{t-1} - \mu) + \cdots + a_p(Z_{t-p} - \mu) + \varepsilon_t \\ &= \sum_{k=1}^p a_k(Z_{t-k} - \mu) + \varepsilon_t. \end{aligned} \quad (4)$$

- $(Z_t)_t$  est dit **moyenne mobile d'ordre  $q$** , noté **MA(q)**, si

$$\begin{aligned} Z_t &= \mu + \varepsilon_t + b_1 \varepsilon_{t-1} + \cdots + b_q \varepsilon_{t-q} \\ &= \mu + \sum_{k=0}^q b_k \varepsilon_{t-k}, \text{ où } b_0 = 1. \end{aligned}$$

## Stationnarité d'un processus ARMA

### Ecriture simplifiée

#### Definition

On appelle **opérateur de retard** l'application  $L$  qui fait associer à la valeur présente  $Z_t$  la valeur retardée  $Z_{t-1}$ , c'est à dire que

- $L(Z_t) = Z_{t-1},$
- $L^2(Z_t) = L(L(Z_t)) = L(Z_{t-1}) = Z_{t-2}.$

L'équation (3) peut alors s'écrire sous une forme plus simple.

### Definition

Soit  $(Z_t)_t$  un processus ARMA(p,q) défini par l'équation (3).

Alors

$$A(L)(Z_t - \mu) = B(L)\varepsilon_t, \quad (5)$$

où  $A(z) = 1 - a_1z - \dots - a_pz^p$  et  $B(z) = 1 + b_1z + \dots + b_qz^q$  sont deux polynômes. On suppose généralement que  $A$  et  $B$  n'ont pas de racines communes.

## Propriété de stationnarité

### Theorem

Soit  $(Z_t)_t$  un processus ARMA( $p, q$ ) défini par l'équation (5).  
 $(Z_t)_t$  est stationnaire si et seulement si le polynôme  $A$  n'admet pas de racine unitaire (de module égal à 1).

En particulier, un processus MA(q) est toujours stationnaire.

## Proposition

*Soit  $(Z_t)_t$  un processus ARMA( $p,q$ ) stationnaire centré associé au bruit blanc  $(\varepsilon_t)_t$  de variance  $\sigma_\varepsilon^2$ . Alors  $(Z_t)_t$  peut s'écrire comme un processus linéaire associé à  $(\varepsilon_t)_t$  ;*

$$Z_t = \sum_{k \in \mathbb{Z}} \beta_k \varepsilon_{t-k} \quad \text{avec } \beta_k \text{ tels que}$$

$$C(z) = \frac{B(z)}{A(z)} = \sum_{k \in \mathbb{Z}} \beta_k z^k.$$

## Mémoire d'un processus ARMA

### Processus MA

Soit  $(Z_t)_t$  un processus MA(1) centré associé au bruit blanc  $(\varepsilon_t)_t$  de variance  $\sigma^2$  :

$$Z_t = \varepsilon_t + b_1 \varepsilon_{t-1}.$$

Alors

$$\begin{cases} \gamma(0) &= \text{var}(Z_t) = \sigma^2(1 + b_1^2), \\ \gamma(-1) &= \gamma(1) = b_1 \sigma^2, \\ \gamma(h) &= 0 \text{ pour tout } |h| \geq 2. \end{cases}$$

Donc la fonction d'autocorrélation est donnée par :

$$\begin{cases} \rho(0) = 1, \\ \rho(-1) = \rho(1) = \frac{b_1}{1+b_1^2}, \\ \rho(h) = 0 \text{ pour tout } |h| \geq 2. \end{cases}$$

Ainsi la mémoire d'un processus MA(1) est finie et s'annule à partir de l'ordre  $h = 2$ .

Soit  $(Z_t)_t$  un MA(q) centré. Alors,

$$\begin{aligned}\gamma(h) &= E(Z_{t+h} Z_t) = 0 \text{ si } |h| > q, \\ &= \sum_{l=0}^q \sum_{m=0}^q b_l b_m E(\varepsilon_{t+h-l} \varepsilon_{t-m}), \text{ si } |h| \leq q, \\ &= \sigma^2 \sum_{j=0}^{q-h} b_{h+j} b_j, \text{ si } |h| \leq q.\end{aligned}$$

## Theorem

Soit  $(Z_t)_t$  un  $MA(q)$ . Alors  $\gamma(q) \neq 0$  et  $\gamma(h) = 0, \forall |h| > q$ .

Soit  $(Z_t)_t$  un  $MA(\infty)$  centré, c'est dire que

$$Z_t = \sum_{k=0}^{\infty} b_k \varepsilon_{t-k}.$$

Alors

$$\gamma(h) = \sigma^2 \sum_{j=0}^{\infty} b_j b_{j+|h|}, \quad (6)$$

$$\text{et } \rho(h) = \sum_{j=0}^{\infty} \frac{b_j b_{j+|h|}}{\sum_{k=0}^{\infty} b_k^2}. \quad (7)$$

Remarque : La fonction d'autocorrélation (ou la mémoire) d'un processus MA( $\infty$ ) ne dépend pas de la variance  $\sigma^2$  du bruit  $(\varepsilon_t)_t$ .

## Processus AR

Soit  $(Z_t)_t$  un AR(1) centré :

$$Z_t = a_1 Z_{t-1} + \varepsilon_t, \text{ avec } |a_1| < 1 \text{ et } a_1 \neq 0.$$

Première méthode pour calculer la mémoire du processus AR(1)

On a

$$\begin{aligned} Z_t &= a_1 Z_{t-1} + \varepsilon_t \\ &= a_1 (a_1 Z_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &\vdots \\ &= \sum_{j=0}^{\infty} a_1^j \varepsilon_{t-j}. \end{aligned} \tag{8}$$

La représentation (8) peut aussi se déduire à l'aide de l'opérateur retard. En effet, en posant  $A(L) = I - a_1 L$ , on obtient

$$Z_t = a_1 Z_{t-1} + \varepsilon_t \iff A(L)Z_t = \varepsilon_t.$$

Or, si  $|a_1| < 1$ , on a

$$(1 - a_1 L)(1 + a_1 L + a_1^2 L^2 + \dots) = 1,$$

donc l'opérateur  $(1 - a_1 L)$  est inversible d'inverse

$$(1 - a_1 L)^{-1} = 1 + \sum_{j=1}^{\infty} a_1^j L^j.$$

Par conséquent,

$$A(L)Z_t = \varepsilon_t \iff Z_t = (1 - a_1 L)^{-1} \varepsilon_t = \sum_{j=0}^{\infty} a_1^j \varepsilon_{t-j}.$$

Donc un  $AR(1)$  est un processus  $MA(\infty)$  et sa fonction d'autocorrélation peut s'obtenir de l'équation (7) avec  $b_j = a_1^j$ .

*Seconde méthode pour calculer la mémoire du processus AR(1)*

De l'égalité (8) on déduit que pour tout

$h > 0$ ,  $E(\varepsilon_{t+h} Z_t) = 0$ . Donc

$$\begin{aligned}
 \gamma(h) &= E(Z_{t+h} Z_t) \\
 &= E((a_1 Z_{t+h-1} + \varepsilon_{t+h}) Z_t) \\
 &= a_1 E(Z_{t+h-1} Z_t) \\
 &= a_1 \gamma(h-1) \\
 &\quad \vdots \\
 &= a_1^h \gamma(0). \tag{9}
 \end{aligned}$$

Donc la fonction d'autocorrélation est donnée par :

$$\begin{aligned}\rho(h) &= a_1^h \\ &= \{signe(a_1)\}^h e^{h \ln(|a_1|)},\end{aligned}\tag{10}$$

où

$$signe(a_1) = \begin{cases} 1 & \text{si } a_1 > 0 \\ -1 & \text{si } a_1 < 0. \end{cases}$$

Ainsi la mémoire d'un processus AR(1) est infinie mais tend vers 0 avec une vitesse exponentielle.

Soit  $(Z_t)_t$  un AR( $p$ ) centré.

*Première méthode pour calculer la mémoire du processus AR( $p$ )*

---

On pose  $A(z) = 1 - a_1z - \dots - a_pz^p$ , alors

$$A(L)Z_t = \varepsilon_t .$$

On peut calculer sa mémoire en écrivant  $A^{-1}(z)$  sous la forme

$A^{-1}(z) = \sum_{k=0}^{\infty} \psi_k z^k$ . On a alors

$$Z_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} , \quad (11)$$

et on applique l'Equation (6).

*Seconde méthode pour calculer la mémoire du processus AR( $p$ )*

La deuxième méthode consiste à calculer la fonction d'autocovariance récursivement. En effet, de l'équation (11) on déduit que pour tout  $k \geq p$ , on a  $E(\varepsilon_{t+k} Z_t) = 0$  et donc

$$\begin{aligned}\gamma(k) &= E(Z_{t+k} Z_t) \\ &= E((a_1 Z_{t+k-1} + \dots + a_p Z_{t+k-p} + \varepsilon_{t+k})Z_t) \\ &= a_1 \gamma(k-1) + \dots + a_p \gamma(k-p).\end{aligned}\quad (12)$$

En comparant les Equations (4) et (12), on constate que la fonction d'autocovariance vérifie la même équation de récursion que le modèle AR( $p$ ) dans laquelle on supprime le bruit.

## Processus ARMA

Si  $(Z_t)_t$  est un ARMA(p,q), on peut montrer que la fonction d'autocovariance vérifie

$$\gamma(k) = a_1\gamma(k-1) + \dots + a_p\gamma(k-p) + \sigma^2 \sum_{k \leq j \leq q} b_j a_{j-k}$$

si  $0 \leq k < \max(p, q + 1)$  et

$$\gamma(k) = a_1\gamma(k-1) + \dots + a_p\gamma(k-p)$$

si  $k \geq \max(p, q + 1)$ .

## Représentation canonique d'un processus ARMA

### Definition

Soit  $(Z_t)_t$  le processus ARMA défini par l'équation (5). On dit que  $(Z_t)_t$  est

- **causal** si les racines de  $A(z)$  sont toutes strictement à l'extérieur du disque unité.
- **inversible** si les racines de  $B(z)$  sont toutes strictement à l'extérieur du disque unité.

## Theorem

*Soit  $(Z_t)_t$  le processus ARMA défini par l'Equation (5), tel que A et B n'ont pas de racines de module égal à 1. Alors il existe deux polynômes  $\tilde{A}(z)$  et  $\tilde{B}(z)$  n'ayant que des racines strictement à l'extérieur du disque unité et un bruit blanc faible  $(\eta_t)_t$  tels que:*

$$\tilde{A}(L)(Z_t - \mu) = \tilde{B}(L)\eta_t.$$

*Cette représentation est unique et dite canonique.*

On note  $|z|$  le module du complexe  $z$ , et  $\bar{z}$  son conjugué. Alors on peut réécrire les polynômes  $A$  et  $B$  de la façon suivante

$$A(z) = \prod_{j=1}^p (1 - \alpha_j z) \quad \text{et} \quad B(z) = \prod_{j=1}^q (1 - \beta_j z) \quad \text{avec}$$

$$|\alpha_j| < 1, \text{ si } 1 \leq j \leq r ; \quad |\beta_j| < 1, \text{ si } 1 \leq j \leq s ;$$

$$|\alpha_j| > 1, \text{ si } r + 1 \leq j \leq p ; \quad |\beta_j| > 1, \text{ si } s + 1 \leq j \leq q .$$

Il suffit ensuite de poser

$$\tilde{A}(z) = \prod_{j=1}^r (1 - \alpha_j z) \prod_{j=r+1}^p \left(1 - \frac{1}{\bar{\alpha}_j} z\right);$$

$$\tilde{B}(z) = \prod_{j=1}^s (1 - \beta_j z) \prod_{j=s+1}^q \left(1 - \frac{1}{\bar{\beta}_j} z\right);$$

$$\text{et } \sigma_\eta^2 = \sigma_\varepsilon^2 \frac{\prod_{j=s+1}^q |\beta_j|^2}{\prod_{j=r+1}^p |\alpha_j|^2}.$$

Les processus ARMA sont destinés à la modélisation des séries chronologiques stationnaires, c'est à dire des séries dont la trajectoire fluctue autour d'une constante avec une variance stable dans le temps.

La modélisation ARMA se fait en cinq étapes :

- ① Identification des degrés ( $p, q$ )
- ② Estimation des paramètres  $a_i, b_j$ .
- ③ Réduction du modèle (significativité des paramètres)
- ④ Validation du modèle (tests de diagnostic des résidus)
- ⑤ Sélection du modèle (choix du modèle optimal parmi plusieurs proposés)

# Contenu

## 1 Processus stationnaires au second ordre

- Processus strictement stationnaire
- Processus faiblement stationnaire

## 2 Processus linéaires

- Test KPSS de stationnarité
- Processus ARMA

## 3 Modélisation d'une série chronologique par un processus ARMA

- Identification du processus ARMA

Plusieurs méthodes ont été élaborées pour déterminer les ordres  $p$  et  $q$  du processus ARMA( $p, q$ ) capable de bien décrire la dynamique de la série chronologique.

**Identification d'un MA(q)** Soit  $(Z_t)_t$  un MA( $q$ ) centré.  
Alors la corrélation vérifie

$$\rho(h) = 0 \text{ si } |h| > q,$$

**Identification d'un AR(p)** Soit  $(Z_t)_t$  un AR( $p$ ) centré.  
Alors la corrélation partielle vérifie

$$r(h) = 0 \text{ si } |h| > p,$$

## Identification d'un ARMA(p,q)

### Méthode 1. Corrélation étendue

#### Proposition

Soit  $(Z_t)_t$  un processus ARMA( $p,q$ ). Alors, asymptotiquement,

$$\begin{cases} r_j(k) = 0 & \text{si } j - q > k - p \geq 0 \\ r_j(k) = c(k - p, j - q) & \text{si } 0 \leq j - q \leq k - p, \end{cases}$$

où  $c(k - p, j - q)$  est soit une constante non nulle, soit une variable aléatoire continue à support dans  $]-1, 1[$ .

Table 1: Fonction d'autocorrélation étendue d'un ARMA(1,1)

AR	MA	0	1	2	3	4	5
0		x	x	x	x	x	x
1		x	0	0	0	0	0
2		x	x	0	0	0	0
3		x	x	x	0	0	0
4		x	x	x	x	0	0
5		x	x	x	x	x	0

## Méthode 2. Critères d'information

### Definition

Soit  $L$  la vraisemblance du modèle. On définit

#### Critère d'Akaike (AIC)

$$AIC(p, q) = -2 \ln(L) + 2(p + q); \quad (13)$$

#### Critère Bayésien (BIC) ou (SBC)

$$BIC(p, q) = -2 \ln(L) + 2(p + q) \ln(n); \quad (14)$$

On choisit alors le couple  $(p^*, q^*)$  qui minimise l'un des deux critères.

# Contenu

## 1 Processus stationnaires au second ordre

- Processus strictement stationnaire
- Processus faiblement stationnaire

## 2 Processus linéaires

- Test KPSS de stationnarité
- Processus ARMA

## 3 Modélisation d'une série chronologique par un processus ARMA

- Identification du processus ARMA

Soit  $(Z_t)_t$  un processus ARMA(p,q) centré

$$Z_t = \sum_{k=1}^p a_k Z_{t-k} + \sum_{j=0}^q b_j \varepsilon_{t-j}, \text{ avec } (\varepsilon_t) \sim BB(0, \sigma^2). \quad (15)$$

où  $\theta = (a_1, \dots, a_p, b_1, \dots, b_q)'$  est le vecteur des paramètres inconnus.

Les paramètres  $\theta$  et  $\sigma^2$  peuvent être estimés soit par la méthodes des moindres carrés soit par la méthode du maximum de vraisemblance.

# Contenu

## 1 Processus stationnaires au second ordre

- Processus strictement stationnaire
- Processus faiblement stationnaire

## 2 Processus linéaires

- Test KPSS de stationnarité
- Processus ARMA

## 3 Modélisation d'une série chronologique par un processus ARMA

- Identification du processus ARMA

Notons  $\beta = (\beta_1, \dots, \beta_{p+q})$ , avec  $\beta_i = a_i$  pour  $1 \leq i \leq p$  et  $\beta_i = b_{i-p}$  pour  $p + 1 \leq i \leq p + q$ . Pour tester

$$(\underline{H_0}) : \beta_i = 0 \text{ contre } (\underline{H_1}) : \beta_i \neq 0 ,$$

Soit  $\hat{\sigma}_i^2$  un estimateur de  $\text{var}(\hat{\beta}_i)$  alors on a sous  $(\underline{H_0})$

$$\frac{\hat{\beta}_i}{\hat{\sigma}_i} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

Pour un risque d'erreur  $\alpha$ , si  $\beta_i = 0$  et si  $n$  est assez grand,  $\frac{\hat{\beta}_i}{\hat{\sigma}_i}$  doit appartenir à l'intervalle  $[-u_{1-\alpha/2}, u_{1-\alpha/2}]$ , où  $u_{1-\alpha/2}$  désigne le quantile d'ordre  $1 - \alpha/2$ , c'est à dire que  $P(Z \leq u_{1-\alpha/2}) = 1 - \alpha/2$ , et  $Z$  suit une loi gaussienne standard.

# Contenu

## 1 Processus stationnaires au second ordre

- Processus strictement stationnaire
- Processus faiblement stationnaire

## 2 Processus linéaires

- Test KPSS de stationnarité
- Processus ARMA

## 3 Modélisation d'une série chronologique par un processus ARMA

- Identification du processus ARMA

## Tests de diagnostic des résidus.

Le modèle ajusté à la série chronologique sera jugé valide (on dit aussi adéquat) si la suite des résidus  $(\hat{\varepsilon}_t)_t$  s'approche d'un bruit blanc.

Donc l'analyse des résidus permettra de valider le modèle ou non.

Plusieurs tests d'adéquation existent.

Il est indispensable que les résidus franchissent aussi le test d'autocorrélation.

Si les résidus sont corrélés la modélisation proposée ne sera pas validée.

## Test du bruit blanc.

On désire effectuer le test

$$\begin{cases} (\underline{H_0}) : \varepsilon_t \text{ est un bruit blanc} \\ (\underline{H_1}) : \varepsilon_t \text{ n'est pas un bruit blanc} \end{cases}$$

Deux statistiques sont souvent considérées.

- Statistique de Box-Pierce (1970):

$$Q_{BP}(m) = n \sum_{j=1}^m \hat{\rho}_\varepsilon^2(j),$$

- Statistique Test de Ljung-Box (1978):

$$Q_{LB}(m) = n(n+2) \sum_{j=1}^m \hat{\rho}_\varepsilon^2(j)/(n-m).$$

On rappelle que  $\hat{\rho}_\varepsilon(j)$  est l'autocorrélation empirique de  $\varepsilon_t$  donnée par

$$\hat{\rho}_\varepsilon(j) = \frac{\hat{\gamma}_\varepsilon(j)}{\hat{\gamma}_\varepsilon(0)}, \hat{\gamma}_\varepsilon(j) = \frac{1}{n} \sum_{t=1}^{n-j} (\varepsilon_{t+j} - \bar{\varepsilon})(\varepsilon_t - \bar{\varepsilon}), \bar{\varepsilon} = \frac{1}{n} \sum_{t=1}^n \varepsilon_t.$$

Sous  $(H_0)$ , les deux statistiques  $Q_{BP}(m)$  et  $Q_{LB}(m)$  suivent asymptotiquement une loi  $\chi^2(m)$ . D'après Box et Pierce, il convient de choisir  $m = n/3$ .

Il existe d'autres tests de validation spécifiques aux modèles ARMA(p,q)

## Tests d'autocorrélation du Portemanteau

On désire effectuer le test

$$\left\{ \begin{array}{l} (\underline{H_0}) : Z_t \text{ est un } ARMA(p,q) \\ (\underline{H_1}) : Z_t \text{ n'est pas un } ARMA(p,q) \end{array} \right.$$

On utilise les statistiques de tests mais  $Q_{BP}(m)$  et  $Q_{LB}(m)$  en remplaçant  $u_t$  par  $Z_t$ . Sous  $(\underline{H_0})$ , les deux statistiques  $Q_{BP}(m)$  et  $Q_{LB}(m)$  suivent asymptotiquement une loi  $\chi^2(m - p - q)$ .

Une fois le modèle validé, on peut ensuite réaliser des tests de normalité.

Si la normalité des résidus est rejetée, cela ne remet pas en cause le modèle estimé.

En effet, grâce à la loi forte des grands nombres et le théorème central limite, la consistance ainsi que la normalité asymptotique des estimateurs restent valables si la taille  $n$  de l'échantillon est suffisamment grande. En outre, la normalité des résidus permettra d'ajouter facilement des intervalles de confiance autour des prévisions.

# Contenu

## 1 Processus stationnaires au second ordre

- Processus strictement stationnaire
- Processus faiblement stationnaire

## 2 Processus linéaires

- Test KPSS de stationnarité
- Processus ARMA

## 3 Modélisation d'une série chronologique par un processus ARMA

- Identification du processus ARMA

Pour modéliser des données, plusieurs modèles peuvent être utilisés: des modèles de régression, des modèles ARMA( $p,q$ ),....

Pour sélectionner un modèle optimal parmi plusieurs proposés on peut adopter l'une des deux stratégies:

## Stratégie 1.

On sélectionne le modèle qui minimise l'un des critères d'information, par exemple le critère AIC.

Si les modèles sont emboîtés, on peut réaliser un test de contraintes du rapport de vraisemblance.

## Stratégie 2.

On utilise des critères pour mesurer la qualité d'ajustement des modèles, c'est-à dire la proximité des données estimées par le modèle  $\hat{z}_1, \dots, \hat{z}_n$  avec les données réelles  $z_1, \dots, z_n$ . On choisit le modèle qui minimise des critères suivants :

## ● Biais "Mean Error" (ME)

$$ME = \frac{1}{n} \sum_{k=1}^n z_k - \hat{z}_k .$$

## ● Critère quadratique "Root Mean Square Error" (RMSE)

$$RMSE = \sqrt{MSE}, \quad MSE = \frac{1}{n} \sum_{k=1}^n (z_k - \hat{z}_k)^2 .$$

## ➊ Biais absolu "Mean Absolute Error" (MAE)

$$MAE = \frac{1}{n} \sum_{k=1}^n |z_k - \hat{z}_k|.$$

## ➋ Critère du pourcentage d'erreur moyenne "Mean Percentage Error" (MPE)

$$MPE = \frac{100}{n} \sum_{k=1}^n \frac{z_k - \hat{z}_k}{z_k}.$$

● Critère du pourcentage d'erreur moyenne absolue  
**"Mean Absolute Percentage Error" (MAPE)**

$$MAPE = \frac{100}{n} \sum_{k=1}^n \frac{|z_k - \hat{z}_k|}{|z_k|}.$$

● Critère du pourcentage d'erreur moyenne absolue  
**"Mean Absolute Scaled Error" (MASE)**

$$MASE = \frac{1}{n} \sum_{k=1}^n \frac{|z_k - \hat{z}_k|}{S_r},$$

avec  $S_r = \frac{1}{m-r} \sum_{k=r+1}^m |z_k - z_{k-r}|,$

où  $r=\text{fréquence}(z)$  dépend de la fréquence des données.

Les critères qui mesurent la qualité d'ajustement ont tendance à diminuer avec la complexité du modèle, et favorisent donc les modèles les plus complexes. Pour utiliser ces critères de façon pertinente, il convient de plutôt de les utiliser pour calculer le pouvoir prédictif des modèles sur un échantillon témoin. On procède donc de la façon suivante :

1) On divise la série en deux sous-séries: la première contient entre 80% et 95% premières observations  $S_1 = (z_1, \dots, z_m)$ , (série d'apprentissage); la deuxième contient entre 20% et 5% dernières observations  $S_2 = (z_{m+1}, \dots, z_n)$  (série témoin).

2) Pour chaque modèle proposé, on calcule les prévisions

$\hat{z}_{m+1/m}, \dots, \hat{z}_{m+N/m}$  des variables aléatoires  $Z_{m+1}, \dots, Z_{m+N}$  et  
on calcule les critères précédents en remplaçant  $\hat{z}_1, \dots, \hat{z}_n$  par  
 $\hat{z}_{m+1/m}, \dots, \hat{z}_{n/m}$  et  $z_1, \dots, z_n$  par  $z_{m+1}, \dots, z_n$ .

3) On sélectionne le modèle qui minimise l'un de ces critères.

Il faut noter que les critères relatifs tels que MPE ou MAPE

peuvent donner des valeurs infinies. Hyndman et al.

conseillent d'utiliser plutôt le critère MASE. Mais il faut

remarquer que les critères MASE et MAE, étant égaux à une

constante près, vont donner la même classification des

modèles sur une série donnée.